On the Influence of Transfer Function Noise on Low Frequency Pressure Matching for Sound Zones

Bo Moller, Martin; Kjaer Nielsen, Jesper; Fernandez Grande, Efren; Krarup Olesen, Soren

Published in:
Proceedings of 2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop

Link to article, DOI:
10.1109/SAM.2018.8448689

Publication date:
2018

Document Version
Peer reviewed version

On the Influence of Transfer Function Noise on Low Frequency Pressure Matching for Sound Zones

1st Martin Bo Møller
Research
Bang & Olufsen
Struer, Denmark
mim@bang-olufsen.dk

2nd Jesper Kjær Nielsen
Audio Analysis Lab, CREATE
Aalborg University
Aalborg, Denmark
jkn@create.aau.dk

3rd Efren Fernandez-Grande
Dept. of Electrical Engineering
Technical University of Denmark
Lyngby, Denmark
efg@elektro.dtu.dk

4th Søren Krarup Olesen
Dept. of Electronic Systems
Aalborg University
Aalborg, Denmark
sko@es.aau.dk

***Abstract*—Controlling the sound field in a room with loudspeakers, requires information about how each loudspeaker radiates sound to the regions of interest. This information can be estimated by in-situ measurements of the transfer functions. However, noise in the estimates can severely deteriorate how well the sound field is controlled. Accurate transfer functions are especially important for sound zones, which rely on destructive interference between multiple loudspeakers. In the presented study, transfer function measurements from eight 10" woofers were repeated 30 times in a room. Pressure matching was used to create sound zones between 20-300 Hz. The performance is evaluated for individual and averaged transfer function measurements. Additionally, the effect of Tikhonov regularization based on the estimated noise in the transfer functions is examined.**

***Index Terms*—Sound Zones, Sound Field Control**

I. Introduction

The purpose of sound zones is to provide individual audio experiences to multiple people in the same room, without the use of headphones [1]. Common approaches, [2]–[4], utilize multiple loudspeakers with controllable input signals to control the sound field at predetermined locations in the room. The simplest sound zone system consists of a bright zone where sound is desired, and a dark zone where it should be suppressed. Using superposition, it is possible to combine multiple such systems to reproduce individual audio in multiple zones with minimal acoustic leakage between the zones. Throughout the audible frequency range the wavelength changes significantly, hence different control strategies are required in different frequency ranges [5]. In this article, the frequency range of interest is 20-300 Hz where the degrees of freedom in the sound field is comparable to the number of available loudspeakers [6].

To control the sound field in the room, knowledge of how the available loudspeakers radiate sound to the bright and dark zone is required. This knowledge is approximated by estimating the transfer functions from each loudspeaker to several positions in each zone. In this work, it is assumed that the sound zones system is a feed-forward system where the transfer functions are estimated in a setup procedure. After the initial setup, the transfer function estimates are not updated. Assuming the transfer functions to be time-invariant is only approximately true as they might change slowly e.g. with changing temperature. The consequence of the transfer functions changing after their initial measurement has been investigated in previous literature, [7]–[10]. Here, the time-varying changes of the transfer functions are disregarded. Instead, it is assumed that the estimation of the transfer functions is corrupted by factors such as acoustic background noise and electrical noise in the measurement system. The purpose of this paper is to examine how such measurement noise can affect the resulting sound field control in an experimental setup.

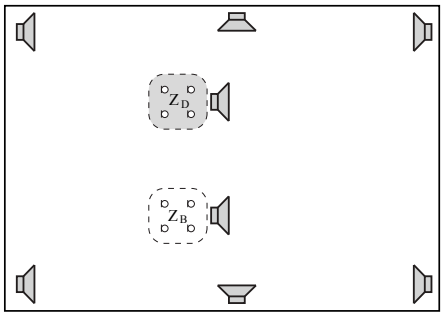

Fig. 1. Sketch of experimental setup, with Z_B and Z_D denoting bright and dark zones, respectively.

II. Theory

A. Pressure Matching

The control method chosen for this investigation is a least squares problem, also referred to as pressure matching in the sound field control literature [11]. The least squares problem to be solved at each angular frequency, ω, is

$$\min_{\mathbf{q}} \|\hat{\mathbf{H}}[\omega]\mathbf{q}[\omega] - \mathbf{d}[\omega]\|_2^2, \tag{1}$$

where $\hat{\mathbf{H}}[\omega] \in \mathbb{C}^{M\times L}$ are the estimated transfer functions from L loudspeakers to M microphones, $\mathbf{q}[\omega] \in \mathbb{C}^{L\times 1}$ is a vector of complex weights for each loudspeaker, and $\mathbf{d}[\omega] \in \mathbb{C}^{K\times 1}$ is the desired pressure at the microphone positions. The desired sound field is chosen as the estimated transfer function of the loudspeaker closest to the bright zone (see Fig. 1) and zero pressure in the dark zone. As seen later, the estimation error of the transfer functions might constitute a significant part of the estimated transfer function matrix. Thus, the solution

might overfit the noisy transfer function measurements and become unstable. One approach to increase the robustness of the solution is to solve the least squares problem with a constraint on the ℓ_2-norm of the complex weights, $\mathbf{q}[\omega]$, also known as Tikhonov regularization [14],

$$\min_{\mathbf{q}}.\|\hat{\mathbf{H}}[\omega]\mathbf{q}[\omega] - \mathbf{d}[\omega]\|_2^2 + \lambda^2[\omega]\|\mathbf{q}[\omega]\|_2^2. \tag{2}$$

A suitable choice of regularization parameter, $\lambda[\omega] \in \mathbb{R}$, is highly problem dependent [14]. In this paper, the focus is on estimating the noise in the measurements and determining the regularization parameter relative to this, rather than to stabilize potential ill-posedness of the underlying physical system.

B. Estimating the noise

To estimate the perturbation in the transfer function matrix, the first step is to estimate the noise in the measured transfer function between a single loudspeaker and microphone in the room. The signal observed with the microphone, from one of K repeated measurement, can be expressed as

$$\mathbf{y}_k = \mathbf{H}\mathbf{x} + \mathbf{e}_k, \tag{3}$$

where $\mathbf{x} \in \mathbb{R}^{N\times 1}$ is the input signal to the loudspeaker, $\mathbf{H} \in \mathbb{R}^{N'\times N}$ is a convolutional matrix of the true impulse response $\mathbf{h}$, and $\mathbf{e}_k \in \mathbb{R}^{N'\times 1}$ is the noise in the measurement. If $\mathbf{x}$ is a length N periodic sequence and N samples are recorded at each measurement ($N' = N$), $\mathbf{H}$ will be a circulant matrix and the observed signal can be written as [12]

$$\mathbf{y}_k = N^{-1}\mathbf{F}^H \operatorname{diag}(\mathbf{F}\mathbf{h})\,\mathbf{F}\mathbf{x} + \mathbf{e}_k, \tag{4}$$

with $\mathbf{F} \in \mathbb{C}^{N\times N}$ being the Discrete Fourier Transform, DFT, matrix as defined in [12] and $(\cdot)^H$ denoting Hermitian transpose. As $\operatorname{diag}(\mathbf{F}\mathbf{h})\,\mathbf{F}\mathbf{x}$ is the entry wise multiplication of the two vectors $\mathbf{F}\mathbf{h}$ and $\mathbf{F}\mathbf{x}$, their order can be interchanged to write the observation vector as

$$\mathbf{y}_k = N^{-1}\mathbf{F}^H \operatorname{diag}(\mathbf{F}\mathbf{x})\,\mathbf{F}\mathbf{h} + \mathbf{e}_k. \tag{5}$$

As the rest of the investigations in this paper are formulated in the frequency domain, both sides of the equation are multiplied by the DFT matrix,

$$\mathbf{F}\mathbf{y}_k = \operatorname{diag}(\mathbf{F}\mathbf{x})\,\mathbf{F}\mathbf{h} + \mathbf{F}\mathbf{e}_k. \tag{6}$$

This is seen to be a linear model estimation problem where $\operatorname{diag}(\mathbf{F}\mathbf{x})$ and $\mathbf{F}\mathbf{y}_k$ are known, and the elements of $\mathbf{F}\mathbf{h}$ are the parameters to be determined. The noise is assumed to be colored Gaussian noise, $\mathbf{e}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$, with unknown covariance matrix $\mathbf{C} \in \mathbb{R}^{N\times N}$. According to [13], if the data segment length, N, is much longer than the autocorrelation time of the colored noise process, the frequency components of the noise samples are asymptotically distributed as

$$E[\omega, k] \overset{a}{\sim} \mathcal{CN}(0, NP_{ee}[\omega]). \tag{7}$$

Here, $E[\omega, k]$ is the entry of $\mathbf{F}\mathbf{e}_k$ corresponding to angular frequency ω and $P_{ee}[\omega]$ is the power spectral density of the noise process evaluated at ω. Each independent frequency component is expressed as $Y[\omega, k] = X[\omega]H[\omega] + E[\omega, k]$ and asymptotically distributed as $Y[\omega, k] \overset{a}{\sim} \mathcal{CN}(X[\omega]H[\omega], NP_{ee}[\omega])$. This is a linear model with the minimum variance unbiased estimator given by [13]

$$\hat{H}[\omega, k] = \frac{X^*[\omega]Y[\omega, k]}{|X[\omega]|^2}. \tag{8}$$

The estimate is asymptotically distributed as $\hat{H}[\omega, k] \overset{a}{\sim} \mathcal{CN}(H[\omega], NP_{ee}[\omega]|X[\omega]|^{-2})$. Combining the information from the K repeated measurements, the approximate log-likelihood function can be written as

$$\ln p(\hat{H}[\omega, \mathbf{k}]; H[\omega], \sigma^2[\omega]) \approx \tag{9}$$
$$-K\ln(\pi) - K\ln(\sigma^2[\omega]) - \frac{1}{\sigma^2[\omega]}\sum_{k=0}^{K-1}(\hat{H}[\omega, k] - H[\omega])^2.$$

In the above equation, $\sigma^2[\omega] = NP_{ee}[\omega]|X[\omega]|^{-2}$ is introduced for compact notation. The asymptotic maximum likelihood estimator for $H[\omega]$ can be determined by taking the partial derivative with respect to $H[\omega]$ and equating the expression to zero, yielding

$$\hat{H}[\omega] = \frac{X^*[\omega]}{K|X[\omega]|^2}\sum_{k=0}^{K-1} Y[\omega, k]. \tag{10}$$

Substituting this estimate of the mean into (9) and determining the stationary point with respect to $\sigma^2[\omega]$, the asymptotic maximum likelihood estimate for the variance is

$$\hat{\sigma^2}[\omega] = \frac{1}{K}\sum_{k=0}^{K-1}(\hat{H}[\omega, k] - \hat{H}[\omega])^2. \tag{11}$$

To evaluate the influence of the noise in the measurement, a confidence region can be determined around the estimate $\hat{H}[\omega]$. The estimate is a linear transformation of a complex Gaussian vector and is thus asymptotically distributed as

$$\hat{H}[\omega] \overset{a}{\sim} \mathcal{CN}\left(H[\omega], \frac{\sigma^2[\omega]}{K}\right). \tag{12}$$

The complex normal distribution can be viewed as consisting of two uncorrelated equal variance normal distributions representing the real and imaginary parts of $H[\omega]$. The confidence region can be described in terms of a circle in the complex plane. The radius of the region is determined as the 95% cumulative probability of the Rayleigh distribution. The confidence region depends on the variance in (12), which is not available. Therefore, the estimate from (11) is used in place of the true variance yielding the estimated radius of the 95% confidence region as

$$\hat{r}_{95\%}[\omega] = \sqrt{-\hat{\sigma^2}[\omega]/K\ln(1 - 0.95)}. \tag{13}$$

C. Transfer function matrix perturbation

The least squares problem in (1) depends on the estimated transfer functions, which generally contain noise. Writing (1) explicitly in terms of the true transfer functions and the perturbation it becomes

$$\min_{\mathbf{q}}.||(\mathbf{H}[\omega] + \Delta\mathbf{H}[\omega])\mathbf{q}[\omega] - \mathbf{d}[\omega]||_2^2, \tag{14}$$

where $\Delta\mathbf{H}[\omega]$ is the perturbation of the true transfer function matrix, $\mathbf{H}[\omega]$, due to the noise in the measurements. Each element in $\mathbf{H}[\omega]$ can be estimated as in (10). It is assumed that the least squares problem is overdetermined, $M > L$. In the case of measurements with poor signal to noise ratio, the largest singular value of the perturbation matrix, $\Delta\mathbf{H}[\omega]$, is significant relative to the smallest singular values of the transfer function matrix, $\mathbf{H}[\omega]$. Hereby, solving (1) would lead to a solution minimizing the residual of transfer functions with significant noise (overfitting). Note that the estimated transfer function matrix, $\hat{\mathbf{H}}[\omega]$, might not be ill-conditioned due to the addition of random perturbation from the measurement noise. Nevertheless, it is of interest to include the estimated knowledge of the perturbation to avoid overfitting. One approach is to add Tikhonov regularization to reduce the influence of singular values of $\mathbf{H}[\omega]$ and $\Delta\mathbf{H}[\omega]$ which are smaller than the largest singular value of $\Delta\mathbf{H}[\omega]$ [14]. To this effect, the regularization parameter $\lambda[\omega]$ in (2) should be chosen to be larger than the largest singular value of the perturbation.

As the true transfer functions are not available, the choice of $\lambda[\omega]$ relies on estimates of both the transfer functions and their variances. In this paper, it is suggested that the perturbation matrix entries are conservatively estimated as the radius of the 95% confidence region around each transfer function, (13). The regularization parameter is chosen as the largest singular value of this matrix (equal to the ℓ_2-norm of the matrix),

$$\lambda[\omega] = \|\hat{\mathbf{R}}_{95\%}[\omega]\|_2. \tag{15}$$

Each $[m, l]$ entry in $\hat{\mathbf{R}}_{95\%}[\omega] \in \mathbb{R}^{M\times L}$ is the 95% confidence radius $\hat{r}_{95\%}[\omega]$ corresponding to the transfer function estimated from loudspeaker l to microphone m at angular frequency ω.

III. Measurements Setup

The experimental investigations for the study were conducted in a 5.66 m by 8.61 m rectangular room with raised ceiling between 2.61 m and 3.71 m, furnished in imitation of a living room. The volume of the room is 143 m^3 and the reverberation time in the investigated frequency range is 0.7 s. The eight woofers were distributed as illustrated in Fig. 1. Transfer functions were measured from each woofer to 18 microphone positions in each zone, non-uniformly sampling a 0.5 m by 0.5 m zone. The microphones were arranged in two height planes (1.10 m and 1.17 m above the floor), with 9 microphones in each. The measurements were conducted as steady state responses, at 48 kHz sampling frequency, with multi tone excitation signals. The multi tone measurements were conducted as progressive measurements of 12 pure tones with random phase shifts, distributed with approximately 1/24th octave spacing between 20 Hz and 297.5 Hz. Approximately because the frequencies of the pure tones were adjusted to be periodic with the 0.4 s measurement duration. The excitation signal was repeated 33 times and recorded as a continuous signal. To avoid clicks at the onset and offset of the measurement sequence, the first and last repetitions were multiplied with half cosine windows and were excluded from the analysis. Furthermore, as the reverberation time of the room at low frequencies is longer than 0.4 s, the first repetition without window is also excluded to ensure 30 repetitions of steady state response measurements. The remaining data were divided into 30 separate segments and analyzed as individual measurement repetitions.

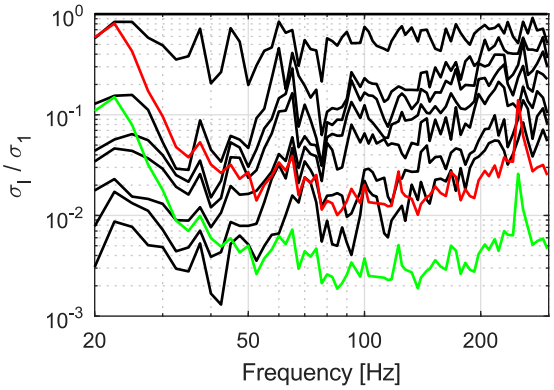

Fig. 2. Black curves: singular values of $\hat{\mathbf{H}}[\omega]$. Red curve: Maximal singular value of $\mathbf{R}_{95\%}[\omega]$ with $K = 1$. Green curve: Maximal singular value of $\mathbf{R}_{95\%}[\omega]$ with $K = 29$. All the curves have been normalized by the maximal singular value of $\hat{\mathbf{H}}[\omega]$ at each frequency.

IV. Results

The first part of the results section is used to summarize the results of the measured transfer functions. The transfer functions and corresponding variances were estimated at the excitation frequencies according to (10) and (11). To evaluate the stability of the least squares problem, the singular values of the estimated matrices $\hat{\mathbf{H}}[\omega]$ are plotted in Fig. 2, normalized to the largest singular value at each frequency. The estimated noise influence is introduced by plotting the largest singular value of $\hat{\mathbf{R}}_{95\%}[\omega]$, normalized to the largest singular value of $\hat{\mathbf{H}}[\omega]$. It is seen, for $K = 1$, that the norm of the estimated perturbation is larger than the smallest singular value of the estimated transfer function below 150 Hz. This indicates that the noise is significant below 150 Hz and that the solution to (1) might overfit the data in this frequency range. For $K = 29$ this is the case below 50 Hz.

In the following, the results of controlling the sound field according to (1) and (2) are presented. As stated in the theory section, the desired pressure in the bright zone was the estimated transfer function from the loudspeaker just behind the zone (see Fig. 1), and the target pressure in the dark zone was zero. The sensitivity to the variations in the estimated transfer functions are investigated with the results plotted in Fig. 3 and 4. Here, the maximum and minimum results across all the measurements combinations are plotted as shaded areas. To relate the results to the common practice of using the transfer functions (either as a single measurement or as an average across multiple measurements), four different evaluations are displayed as areas on the plots.

1) Light gray shaded area: One of the 30 measured transfer functions is used to calculate the loudspeaker weights by solving (1) as normal equations [12], and another

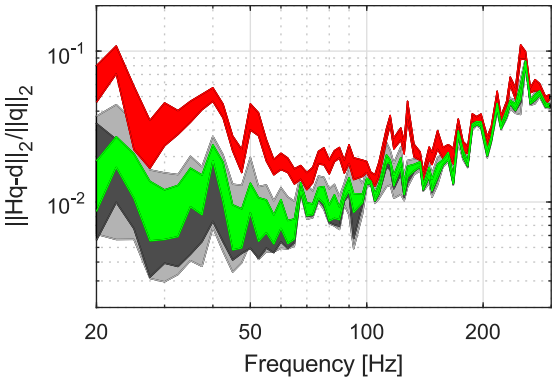

Fig. 3. ℓ_2-norm of the residual error, normalized by the ℓ_2-norm of the loudspeaker weights, calculated at each frequency.

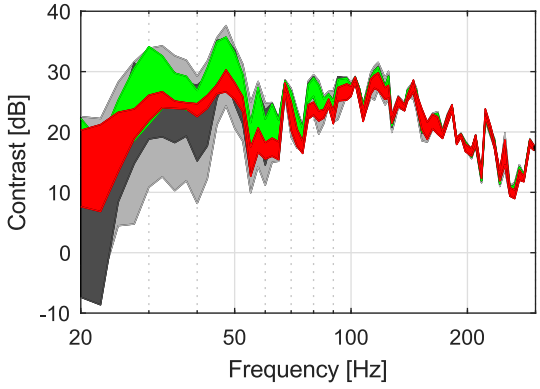

Fig. 4. Ratio of average squared pressure in the bright zone relative to the average squared pressure in the dark zone, displayed in dB.

is used to evaluate the resulting sound field. This is repeated for all (30×29) combinations and the shaded area displays the range between maximum and minimum at each frequency.

2) Dark gray shaded area: 29 of the measurements are averaged and used to calculate the loudspeaker weights by solving (1), and the last measurement is used to evaluate the result. This is repeated for all 30 combinations and the shaded area displays the range between maximum and minimum at each frequency.
3) Red shaded area: The choice of measurements for calculating weights and evaluating the results is identical with 1), however the weights are determined by solving (2). The regularization parameter is chosen according to (15) where the 95% confidence regions has been determined from (15) and (13) with $K = 1$ and the variance was determined according to (11) from all 30 measurements.
4) Green shaded area: The choice of measurements for calculating weights and evaluating the results is identical with 2), however the weights are determined by solving (2). The regularization parameter is given by (15) using (13) with $K = 29$. The variance was determined according to (11) from all 30 measurements. .

Fig. 3 depicts the normalized residual error. It is seen that the regularized results from a single measurement yield a higher residual error than the regularized result from 29 measurements. This is due to the reduced regularization added to the 29-measurement results. However, it is also seen that both regularized results reduce the variation in residual error at low frequencies, relative to the results without regularization.

The contrast (ratio of mean square pressure in the bright and dark zone) results in Fig. 4 show a large variation at low frequencies. It is observed that the regularization reduces the variation in the results and increases the overall contrast. The result with $K = 29$ and regularization, attains similar maximum contrast as the result with $K = 1$ and no regularization. Finally, the regularized solution with $K = 1$ exhibit the lowest variation but does not attain similar maximum contrast due to the increased regularization.

V. Discussion

From the contrast results (Fig. 4), it is clearly seen that there are significant differences between solving (1) and (2). From Fig. 2, it is seen that the estimated perturbation is larger than the smallest singular value of $\hat{\mathbf{H}}[\omega]$ at frequencies below 150 Hz for $K = 1$ and below 50 Hz for $K = 29$. This indicates that solving (1) is overfitting the noisy measurements, which seems to be in agreement with the results. Observing the singular values of $\hat{\mathbf{H}}[\omega]$ in Fig. 2, it is seen that the averaged matrices are ill-conditioned relative to the rule-of-thumb in [16] (the condition number of $\hat{\mathbf{H}}^H[\omega]\hat{\mathbf{H}}[\omega]$ being between 1000 and 5000). This rule-of-thumb is only fulfilled above 150 Hz, which supports the observation that regularization is required below this frequency.

The results presented here highlight a potential challenge in creating sound zones. If possible, the influence of the noise should be minimized by the design of the measurements. However, as conditions such as background noise might not be controllable, estimation of the noise in the measurement is required to ensure a robust solution. It is seen that it is more effective to use a single transfer function measurement and regularization based on the estimated noise than averaging several measurements without regularization. Ideally, one would conduct several transfer function measurements to improve the signal to noise ratio and have a good estimate of the noise variance for determining the regularization parameter.

VI. Conclusion

A measurement study has been conducted to investigate the potential influence of measurement noise in determining loudspeaker weights for creating sound zones. The average transfer functions and noise were estimated from 30 repeated measurements. It was observed that this number of averages was not sufficient to eliminate the influence of the noise. For the given experiment, introducing regularization based on the estimated noise can increase the robustness without significantly reducing the performance. Furthermore, it was seen that a single measurement with regularization based on the estimated noise can provide more stable results than 30 averages without added regularization.

REFERENCES

[1] T. Betlehem, W. Zhang, M. A. Poletti, and T. D. Abhayapala "Personal sound zones," IEEE Signal Processing Magazine, pp. 81–91, March 2015.

[2] J. Choi and Y. Kim "Generation of an acoustically bright zone with an illuminated region using multiple sources," J. Acoust. Soc. Am, Vol. 111, No. 4, pp. 1695–1700, April 2002.

[3] P. Coleman, P. J. B. Jackson, M. Olik, M. Møller, M. Olsen, and J. A. Pedersen "Acoustic contrast, planarity and robustness of sound zone methods using a circular loudspeaker array," J. Acoust. Soc. Am., Vol. 135, No. 4, pp. 1929–1940, April 2014.

[4] P. Coleman, P. J. B. Jackson, M. Olik, and J. A. Pedersen "Personal audio with planar bright zone," J. Acoust. Soc. Am., Vol. 136, No. 4, pp. 1725–1735, October 2014.

[5] W. F. Druyvesteyn and J. Garas, "Personal sound," J. Audio Eng. Soc., Vol. 45, No. 9, pp. 685–701, September 1997.

[6] L. Brännmark, A. Bahne, and A. Ahlén, "Compensation of loudspeaker-room responses in a robust mimo control framework," IEEE Transactions on Audio, Speech, and Language Processing, vol. 21, no. 6, October 2013.

[7] J. Park, J. Choi, and Y. Kim "Acoustic contrast sensitivity to transfer function errors in the design of a personal audio system," J. Acoust. Soc. Am., vol. 134, no. 1, pp. EL112–EL118, June 2013.

[8] Y. Cai, L. Liu, M. Wu, and J. Yang, "Robust time-domain acoustic contrast control design under uncertainties in the frequency response of the loudspeakers," Inter-Noise 2014 Proceedings, November 2014.

[9] Q. Zhu, P. Coleman, M. Wu, and J. Yang "Robust Acoustic Contrast Control with Reduced In-situ Measurement by Acoustic Modeling," J. Audio. Eng. Soc., vol. 65, no. 6, pp. 460-473, June 2017.

[10] M. Olsen and M. B. Møller "Sound Zones: On the effect of ambient temperature variations in feed-forward systems," Audio Engineering Society Convention 142, May 2017.

[11] M. Poletti "An investigation of 2d multizone surround sound systems," Audio Engineering Society Convention 125, October 2008.

[12] G. H. Golub and C. F. van Loan, Matrix Computations, 4. edt., Johns Hopkins University Press, 2013.

[13] S. M. Kay, Fundamentals of Statistical Signal Processing: Estimation Theory, Prentice-Hall, 1993.

[14] P. C. Hansen, Discrete Inverse Problems: Insights and Algorithms, SIAM, 2010.

[15] T. Betlehem and C. Withers "Sound field reproduction with energy contraint on loudspeaker weights," IEEE Transactions on Audio, Speech, and Language Processing, vol. 20, no. 8, October 2012.

[16] O. Kirkeby, P. A. Nelson, F. Orduna-Bustamante, and H. Hamada "Local sound field reproduction using digital signal processing," J. Acoust. Soc. Am., vol. 100, no. 3, September 1996.